A NOTE ON LINEAR DISCRIMINANTS VIA IDEMPOTENT MATRICES

by

Seymour Geisser

Technical Report No. 163

November 1971

University of Minnesota
Minneapolis, Minnesota

1. Introduction.

Multiple linear discriminants are obtained in the multivariate normal case from the ratio of likelihoods, e.g., Anderson [1, p. 148]. If no distributional assumptions are made, they can, as Fisher [3, p. 382] originally did, be obtained from his notion of "best linear discriminant"--that linear combination of the components of the vector mean difference between two populations which is a maximum relative to its standard deviation. These may then be taken pairwise over all the mean vectors to generate the space of "best linear discriminants." Discriminants may also be obtained by maximizing some distance or measure of spread amongst the mean vectors relative to the common covariance matrix. The methods employed for the maximization are usually differentiation (though verification that the solution actually leads to a maximum is often omitted Wilks [4, p. 575]) or geometrical arguments, e.g., Dempster [2, p. 220]. We shall present here an alternative algebraic derivation which depends on a simple theorem involving idempotent matrices.

An incidental feature of this approach is the concomitant demonstration of an obvious, but rarely stressed, fact--that the same set of linear discriminants maximizes any measure of spread that is an increasing scalar function of the non-zero roots of the product of the "between matrix" of means and the inverse of the covariance matrix.

2. The Maximization Theorem.

We now prove the following:

Theorem.

Let Z be a real $p \times m$ matrix of rank $s = \min(p, m)$ and E_k be the class of $p \times p$ real symmetric idempotent matrices of rank k. Then for all $F \in E_k$ the maximum attainable values that the non-zero ordered roots a_i of $Z'FZ$ are α_i, $i = 1, \ldots, t$, $t = \min(k, m)$, where α_i are the non-zero ordered roots of $Z'Z$.

<u>Proof</u>:

Let P be the orthogonal matrix such that

$$D_m = P'Z'ZP = P'Z'FZP + P'Z'(I-F)ZP \tag{1}$$

or for $Y = ZP$

$$D_m = Y'Y = Y'FY + Y'(I-F)Y$$

where

$$D_j = \begin{pmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \alpha_j \end{pmatrix}.$$

Hence the roots of $Z'FZ$ are the roots of $Y'FY$ and by virtue of (1) the ordered roots of $Y'FY$ $a_1 \geq \cdots \geq a_m \geq 0$ are less than the ordered roots of $Y'Y$, i.e., $\alpha_i \geq a_i$, $i = 1, \ldots, m$. Now let Q be a $p \times p$ orthogonal matrix such that $Q = (Y_{(s)}D_s^{-\frac{1}{2}}, Q_2)$, where $Y_{(s)}$ consists of the first s columns of Y. Then

$$Y'FY = Y'QQ'FQQ'Y = Y'QF^*Q'Y$$

where F^* is obviously idempotent. Further

$$Y'Q = \begin{array}{c c} & \begin{array}{cc} s & p-s \end{array} \\ \begin{array}{c} s \\ m-s \end{array} & \left(\begin{array}{c|c} D_s^{\frac{1}{2}} & 0 \\ \hline 0 & 0 \end{array}\right) \end{array}$$

so that

$$Y'FY = Y'Q'F^*QY = \begin{array}{c c} & \begin{array}{cc} s & m-s \end{array} \\ \begin{array}{c} s \\ m-s \end{array} & \left(\begin{array}{c|c} D_s^{\frac{1}{2}} F_{11}^* D_s^{\frac{1}{2}} & 0 \\ \hline 0 & 0 \end{array}\right) \end{array} \tag{2}$$

and F_{11}^* is the $s \times s$ matrix in the upper left hand corner of F^*.

Now the rank of $Y'FY$ is $t = \min(k, m)$ or $t = \min(k, s)$. Hence all solutions, for which $a_i = \alpha_i$ $i = 1,\ldots, t$ are such that the first t diagonal elements of F_{11}^* are 1 since $\Sigma_{i=1}^{t} \alpha_i = \Sigma_{i=1}^{t} \alpha_i f_{ii}^*$, $0 \leq f_{ii}^* \leq 1$ and $\Sigma_{i=1}^{p} f_{ii}^* = k$ must be satisfied. This implies that the off diagonal elements in those rows and columns are zero since we are dealing with idempotent matrices. Therefore all solutions for F^* are

$$F_0^* = \begin{array}{c} \\ k \\ p-k \end{array}\!\!\begin{array}{c} \begin{array}{cc} k & p-k \end{array} \\ \left(\begin{array}{c|c} I_k & 0 \\ \hline 0 & 0 \end{array}\right) \end{array} \quad \text{for } t = k, \text{ i.e., } k \leq m$$

or

$$F_0^* = \begin{array}{c} \\ m \\ p-m \end{array}\!\!\begin{array}{c} \begin{array}{cc} m & p-m \end{array} \\ \left(\begin{array}{c|c} I_m & 0 \\ \hline 0 & G \end{array}\right) \end{array} \quad \text{for } t = m, \text{ i.e., } m \leq k,$$

where G is any idempotent $p - m \times p - m$ matrix of rank $k - m$. Hence the totality of solutions for F are $F_0 = QF_0^*Q'$, so that

$$(3) \qquad F_0 = Y_{(k)}D_k^{-1}Y'_{(k)} \quad \text{for } k \leq m$$

which is unique and

$$(4) \qquad F_0 = YD_m^{-1}Y' + Q_2GQ_2' \quad \text{for } m \leq k.$$

Corollary 1.

If $m \leq k$, the totality of solutions for F are

$$(5) \qquad F_0 = Z(Z'Z)^{-1}Z' + E_{k-m}$$

where E_{k-m} is an idempotent matrix of rank $k - m$ orthogonal to Z. Further $Z'Z = Z'F_0Z$ and if $m = k$ the solution $F_0 = Z(Z'Z)^{-1}Z'$ is unique.

Proof:

Note that from (4) and $ZP = Y$ that

$$F_0 = ZPD_m^{-1}P'Z + Q_2GQ_2' = Z(Z'Z)^{-1}Z' + Q_2GQ_2'.$$

Hence set $Q_2GQ_2' = E_{k-m}$ since G is an arbitrary idempotent matrix of rank $k - m$ and Q_2 is orthogonal to Z being it is orthogonal to Y. Multiplication of (5) on the left by Z' and on the right by Z yields $Z'Z = Z'F_0Z$. Since E_0 is necessarily the null matrix the uniqueness part follows immediately.

Corollary 2.

If $g(Z'FZ) = g(a_1, \ldots, a_t)$ is a scalar increasing function of the roots a_i, then

$$\max_{F \in E_k} g(a_1, \ldots, a_t) = g(\alpha_1, \ldots, \alpha_t).$$

The proof of the corollary is an immediate consequence of the theorem and the fact that $a_i \leq \alpha_i$. This includes such functions $\mathrm{Tr}\, Z'FZ$ and $|I + Z'FZ|$, yielding

$$\max_{F \in E_k} \mathrm{Tr}\, Z'FZ = \sum_{i=1}^{t} \alpha_i = \begin{cases} \mathrm{Tr}\, Z'Z & \text{if } m \leq k \\ \sum_{i=1}^{k} \alpha_i & \text{if } k \leq m \end{cases}$$

$$\max_{F \in E_k} |I + Z'FZ| = \prod_{i=1}^{t} (1+\alpha_i) = \begin{cases} |I + Z'Z| & \text{if } m \leq k \\ \prod_{i=1}^{k} (1+\alpha_i) & \text{if } k \leq m. \end{cases}$$

3. Application.

Suppose there are r p-dimensional multivariate populations with means $\mu_1, \ldots, \mu_r$ and common positive definite covariance matrix Σ. Further let $\beta = \sum_{i=1}^{r} (\mu_i - \bar{\mu})(\mu_i - \bar{\mu})'$ and $\bar{\mu} = r^{-1} \sum_{i=1}^{r} \mu_i$ where β is of rank $r - v$. Assume that $g(\Sigma^{-1}\beta)$ is any scalar measure of the spread of these r populations that is increasing in the non-zero roots of $\Sigma^{-1}\beta$,

$\delta_1 \geq \ldots \geq \delta_s > 0$. Suppose further we transform these r p-dimensional populations into a $k \leq p$ space by a real transformation matrix $C_{k \times p}$ which is of rank k. Hence $\eta_i = C\mu_i$, $i = 1, \ldots, r$, $\Omega = C\Sigma C'$ and $\Gamma = C\beta C'$ and the measure of spread in k dimensions is $g_k(\Omega^{-1}\Gamma)$, i.e., the same scalar function of the non-zero roots $d_1 \geq \ldots \geq d_t > 0$ of $\Omega^{-1}\Gamma$ where $t = \min(k, r-v)$. Now the non-zero roots of $\Omega^{-1}\Gamma = (C\Sigma\, C')^{-1}C\beta C$ are the same as the non-zero roots of $\Lambda' C'(C\Sigma\, C')^{-1}C\Lambda$ where $\beta = \Lambda\Lambda'$ and Λ is $p \times r-v$. Set $C\Sigma^{\frac{1}{2}} = H$ where $\Sigma^{\frac{1}{2}}$ is the positive definite symmetric square root of Σ so that the non-zero roots of $\Omega^{-1}\Gamma$ are the same as the non-zero roots of $\Lambda'\Sigma^{-\frac{1}{2}} H'(HH')^{-1}H\Sigma^{-\frac{1}{2}}\Lambda$. Let $r - v = m$ and $Z = \Sigma^{-\frac{1}{2}}\Lambda$ and set the idempotent matrix $H'(HH')^{-1}H = F$. Hence as by our previous corollary

$$\operatorname*{Max}_{C}\, g_k(\Omega^{-1}\Gamma) = \operatorname*{Max}_{F}\, g_k(Z'FZ) = g(\delta_1, \ldots, \delta_t).$$

Certain special cases of interest are the Hotelling measure of spread (based on T_0^2) $H = \operatorname{Tr}\Sigma^{-1}\beta$ in k-dimensions $H_k = \operatorname{Tr}\Omega^{-1}\Gamma$, and similarly a Wilks measure $W = |I + \Sigma^{-1}\beta|$, $W_k = |I + \Omega^{-1}\Gamma|$, based on the inverse of his Λ criterion. Hence

$$\operatorname*{Max}_{C}\, H_k = \sum_{i=1}^{t} \delta_i = \begin{cases} \operatorname{Tr}\Sigma^{-1}\beta & \text{if } m \leq k \\ \sum_{i=1}^{k} \delta_i & \text{if } k \leq m \end{cases}$$

$$\operatorname*{Max}_{C}\, W_k = \prod_{i=1}^{t} (1+\delta_i) = \begin{cases} |I + \Sigma^{-1}\beta| & \text{if } m \leq k \\ \prod_{i=1}^{k} (1+\delta_i) & \text{if } k \leq m. \end{cases}$$

To find solutions for C we note that there is an orthogonal matrix P such that

$$P'\Lambda'\Sigma^{-1}\Lambda P = \Delta_m = \begin{pmatrix} \delta_1 & & 0 \\ & \ddots & \\ 0 & & \delta_m \end{pmatrix}$$

and we set $Y = \Sigma^{-\frac{1}{2}}\Lambda P$ and $Y_{(j)} = \Sigma^{-\frac{1}{2}}\Lambda P_{(j)}$ where $P_{(j)} = (P_1, \ldots, P_j)$ is the matrix consisting of the first j columns of P. Note also that ΛP_i is the invariant vector of $\beta\Sigma^{-1}$ corresponding to the root δ_i, $i = 1, \ldots, m$ and the calculation of Y does not depend on Λ. Hence from the theorem

$$F_0 = Y_{(k)}\Delta_k^{-1}Y'_{(k)} \qquad \text{for } k \leq r - 1$$

$$F_0 = Z(Z'Z)^{-1}Z' + E_{k-m} \qquad \text{for } r - 1 \leq k.$$

From $H'(HH')^{-1}H = F$ we obtain $H = HF$ and noting from (3) that $Y'_{(k)}F_0 = Y'_{(k)}$ then $H_0 = Y'_{(k)}$ and $C_0 = Y'_{(k)}\Sigma^{-\frac{1}{2}}$ for $k \leq m$. For $k \geq m$ $Y'F_0 = Y'$ from (4), leaving open $k - m$ additional orthogonal vectors to complete $H_0 = (Y', Q_2)$ yielding $C_0 = (Y', Q_2)\Sigma^{-\frac{1}{2}}$. We note that in no case is C_0 unique as any solution H_0 can be multiplied by a real non-singular $k \times k$ matrix on the left and still remain a solution because of the invariance of the roots.

For $k = \min(p, r-v) > 1$, we have essentially the multiple discriminant case and for $k = r - v = 1$ it is clear that $C_0 \,\alpha(\mu_1 - \mu_2)\Sigma^{-1}$ which yields Fisher's discriminant. The fraction of total loss sustained in the measure of spread when $k < s$ is

$$L = \frac{g(\delta_1, \ldots, \delta_s) - g(\delta_1, \ldots, \delta_k)}{g(\delta_1, \ldots, \delta_s)}$$

where $s = \min(p, r-v)$. For example if we are using

$$H = \mathrm{Tr}\ \Sigma^{-1}\beta = \sum_{i=1}^{s} \delta_i$$

then

$$L = \sum_{i=k+1}^{s} \delta_i \Big/ \sum_{i=1}^{s} \delta_i \, .$$

It is also clear that when $k > s$, the loss already being 0 when $k = s$, that the additional $k - s$ vectors cannot add anything to the spread. We further note that in most multiple discriminant situations the vectors $\mu_1, \ldots, \mu_r$ are usually linearly independent so that $v = 1$.

REFERENCES

[1] Anderson, T. W. (1958). An Introduction to Multivariate Statistical Analysis. New York: John Wiley and Sons.

[2] Dempster, A. P. (1968). Elements of Continuous Multivariate Analysis. Reading, Massachusetts: Addison-Wesley.

[3] Fisher, R. A. (1938). The statistical utilization of multiple measurements. Annals of Eugenics 8 376-386.

[4] Wilks, S. S. (1962). Mathematical Statistics. New York: John Wiley and Sons.